Cairo University

The 56th Annual International

Conference of Data science

Computer Science

4 – 6, Dec, 2023

Traffic Flow Prediction with Big Data and Machine Learning Techniques

Sayed A. Sayed*, Yasser Abdel-Hamid, and Hesham Ahmed Hefny
Computer Science Department
Faculty of Graduate Studies for Statistical Research
Cairo University, Giza, Egypt
*se.sayedahmed@gmail.com , yasserabdelhamid@gmail.com , hehefny@cu.edu.eg

***Abstract*— The rise of Artificial Intelligence (AI) and Big Data in many industries, including transportation, has inspired new ideas and solutions to complicated problems like traffic congestion, which affects our quality of life. These methods are called Intelligent Transportation Systems (ITS). ITS addresses numerous traffic challenges, including congestion. Additionally, various smart city applications that improve transit and mobility use it. Traffic forecasting is essential in transportation. In addition to route planning and traffic restrictions, it can considerably impact road construction and project design. Therefore, it must be assessed and predicted precisely. Thus, traffic prediction must be accurate and efficient. To increase traffic flow prediction accuracy, new models and frameworks have been rapidly created together with AI and Big Data methods. This study's main goals are: First, to analyze the most frequent traffic prediction machine learning methods. Second, it describes traffic forecast data types. Then, it discusses BD's traffic forecast potential. It concludes by discussing machine learning and big data difficulties and future improvements in traffic prediction.**

Keywords— AI; ITS; Traffic Congestion; Machine Learning; Traffic Prediction

I. INTRODUCTION

As cities have become more crowded and congested over the last few decades, there has been a greater need for the development of ITS-based solutions for accurate traffic prediction and mobility control (Nellore K and Hancke., 2016). ITS is a new way to provide transportation that uses advanced data communication technologies. It does this by combining computers, information technology, communications technologies, and other technologies and using them in the transportation business. The goal of this process is to make a system that works for people, cars, and roads (Patel P, et al., 2019). ITS can be used to build a complete, accurate, real-time, and useful transportation control system (An S, et al., 2011). Besides that, it could cut down on dangers, high accident rates, traffic jams, carbon emissions, and air pollution while also making things safer and more reliable, speeding up travel, easing traffic, and making passengers happier (Qureshi K and Abdullah, 2013).

Accurately predicting traffic flow is an important part of ITS because it helps all the stakeholders involved (passengers, traffic managers, lawmakers, and road users) use transport networks more safely and smartly (Chen C, et al, 2020) (Sun P and Tao, 2020). The quality of the traffic data is what makes these systems work. Only then ITS will be useful. The World Health Organization's (WHO) 2018 report on the universal status of road safety says that road traffic deaths are still going up, with 1.35 million deaths recorded in 2016. This means that studying traffic forecasting is a good way to cut down on traffic and make travel safer and cheaper (Makaba T and Paul, 2020) (World Health Organization, 2018).

In the past, predicting traffic flow relied on parametric models like time series analysis that were based on data from the past. In time series, a set of readings x that were taken at a certain time t are recorded. The goal is to find patterns in past traffic data that show how traffic changed over time and then use these patterns to make predictions. The Kalman Filtering method for time-series analysis (Bengio Y, 2009) was another model for mobile random problems that could solve regression problems and reduce variance to get the best results. Also, the Auto-Regressive Integrated Moving Average (ARIMA) model is a standard and well-known way to guess how traffic will move in the short run (Van.D et al.,1996). There were many changes made to the ARIMA model, and the results showed better performance (Lee S, & Fambro D., 1999 - Williams B., 2001 - Williams B. and Hoel L, 2003 – Chen K,et al., 2020).

Nonparametric models like the Random Forest (RF) Algorithm, the Bayesian Algorithm (BA) method, K-Nearest Neighbor (KNN), Principal Component Analysis (PCA), and Support Vector Algorithms (Bengio Y, 2009) have recently been used to predict traffic flow because it is random and doesn't follow a straight line. Neural networks were also often used to predict how traffic would move (Kashyap A et al., 2022).

There are many areas that have used Big Data (BD) extensively. One of them is the transportation industry. Using a number of different data sources, it is possible to accurately predict and estimate traffic flows, which improves the operation as a whole. Transport is a great example of an area where BD has been used too much. As the Internet of Things (IoT), Cyber-Physical Systems (CPS), and smart cities grew, they made it easier to collect huge amounts of data from things like security cameras, mobile phones, and static sensors. From Trillion bytes to Petabytes (W. Jiang and J., 2022), the amount of data has grown. Several applications consider predicting traffic flow as a major challenge, and that's what this study will focus on. It is sometimes impossible to get traffic

information (like volume, speed, trip time, etc.) without changing the raw data, but it is possible to get other information about traffic conditions.

The main goals of this research are to analyze the most widely used machine learning approaches for traffic flow prediction. In addition, it presents the potential BD role in traffic prediction. Finally, it discusses the challenges facing and potential future developments in machine learning and big data concerning traffic prediction.

Here's how the remainder of this survey is put together: Section 2 talks about how the traffic prediction problem, machine learning, and big data work from a theoretical point of view. Section 3 explains how the survey was done and gives a review of the research on how machine learning and big data were utilized to predict traffic flow. In Section 4, we talk about the problems that already exist in this survey's topic. Section 5 is the last part of the paper which concludes it.

II. BACKGROUND

Many high-resolution traffic data from ITS can be used to estimate traffic flow (J. Zhang et al., 2011). Predicting traffic flow is a time series problem that uses data from past observation stations to estimate future flow counts. Traffic modeling, operation, and management involve anticipating traffic flow. Real-time traffic predictions can help road travelers save money and make better judgments. It can help traffic authorities reduce congestion by improving traffic control. AI branch machine learning has grown rapidly in recent years (Chowdary G, 2021). These approaches accurately forecast traffic. Many excellent traffic data collection systems and Big Data technology for storing and analyzing big data sets have enabled many traffic volume prediction methods.

1. Machine Learning

Machine Learning (ML) approaches are statistical models used to classify and predict outcomes based on input data (Chowdary G, 2021). ML is a branch of AI that concentrates on creating prediction algorithms by uncovering patterns in large datasets, without being tailored for a specific task (Singh G, 2018). ML models are categorized into three groups based on the learning methodologies they utilize: supervised learning, unsupervised learning, and reinforced learning (RL), as depicted in Figure 1.

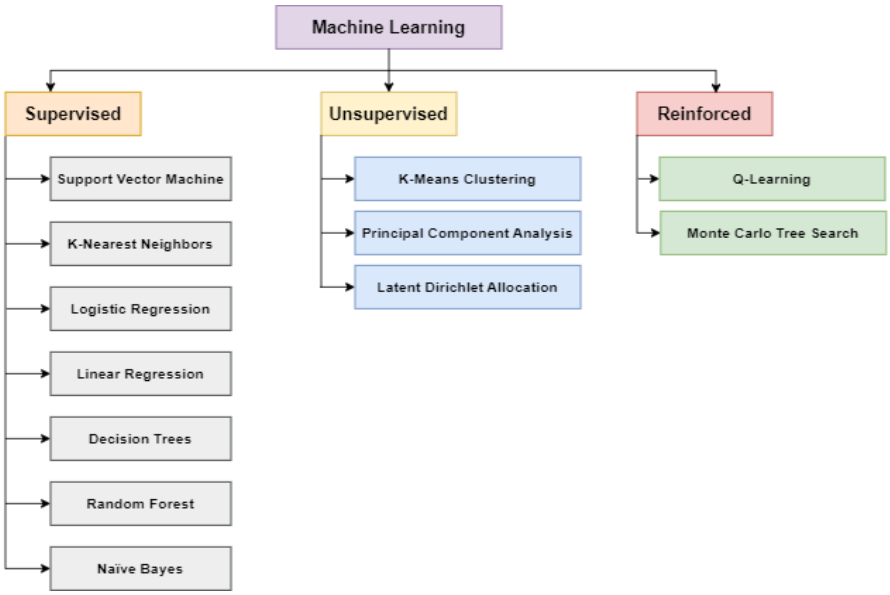

Fig. 1. *Machine Learning Categories.*

1.1 Supervised Learning

In supervised learning tasks, the model is provided with a labeled dataset consisting of feature vectors and their corresponding anticipated output labels. The goal of these models is to establish an inference function that accurately associates feature vectors with corresponding output labels. Upon the completion of ML model training, it is capable of generating predictions using novel data. Supervised learning methods (Chowdary G, 2021) can be used to provide predictions that are either continuous or discrete. Supervised learning algorithms, such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Linear Regression, Decision Trees (DT), Random Forests (RF), and Naive Bayes, are supervised learning approaches examples (Singh G, 2018).

A. Support Vector Machine

SVM is a deterministic linear classifier that focuses on classification challenges. SVM is considered the best ML algorithm. The core of SVM is margin calculation. This method depicts each data point as a point in an n-dimensional space, where n is the number of features, and each feature is a coordinate. This method analyzes vectorized data to create a hyperplane that separates the classes (Dey

A, 2016). Next, several margins are drawn between classes, and a hyperplane is created to reduce mean-squared error and maximize margin-class distance (Dhall D and Juneja M, 2020). After finding an optimal separating hyperplane for linearly separable data, support vector points are the data points on its boundary. A linear combination of these locations is then provided as the solution. We ignore the remaining data (Kotsiantis S and Pintelas P, 2007). Thus, the number of features in the training data does not affect SVM model complexity. SVMs excel at learning problems with many features compared to training examples.

Even though the SVM has a large margin to assess several hyperplanes, misclassified instances may prevent it from finding one that divides the data. A soft margin that allows misclassifications of certain training samples may help (Veropoulos K and Cristianini N, 1999). Only binary classification issues can be solved by SVM. Thus, multi-class problems must be converted into binary classification problems. Analyzing categorical data is tough, although scaling can improve results (Kotsiantis S and Pintelas P, 2007).

B. K-Nearest Neighbors

KNN is a nonparametric classification approach that makes no dataset assumptions. Its efficiency and simplicity are admired. KNN method predicts unlabeled data class using a labeled training dataset. To classify new inputs, KNN is used on clustered datasets. KNN is useful when data is unfamiliar (Taunk, K and Swetapadma A, 2019). KNN uses a range of variable values, usually 0–1, to find the closest training data points. KNN uses Manhattan, Euclidean, Minkowski, and Hamming distances. Nearest neighbors are calculated using Euclidean distance for continuous data. Categorical data use the Hamming distance function (Obulesu O et al., 2018).

Choosing the K value is the hardest part of the KNN algorithm because it affects how well and accurately the algorithm works. When predicting a class label, small K values cause noise, while large K values may cause too much fitting likelihood. It also takes longer to compute and slows down the speed of execution. The K value is estimated according to equation (1):

$$K = n\hat{}(1/2) \qquad (1)$$

Where n represents the size of the dataset.

The training data will undergo cross-validation using different K values to optimize the test results. The optimal value for test results will be determined based on the optimal precision (Obulesu O et al., 2018).

The KNN method is simple and easy to utilize. This classification method is adaptable and ideal for multi-mode classes. The KNN approach for classifying unfamiliar data is expensive. Distance between the k-nearest neighbors must be calculated. As training set size increases, algorithm computations become more difficult. Disruptive or insignificant features reduce precision. KNN keeps all training data without generalizing it. Thus, more data dimensions reduce location accuracy. KNN learns slowly since it calculates k neighbor distances (Ray S., 2019).

C. Logistic Regression

Logistic regression is a supervised learning approach for discriminating among three or more classes (Dey A., 2016). It quantifies the probability of an event occurrence using binary values (0 and 1) and is dependent on the input variables, resulting in a binomial conclusion. An example of a binomial outcome in Logistic Regression is the prediction of whether an email is classified as spam or not. Furthermore, Logistic Regression has the capability to generate ordinal outcomes, such as assigning a rating to a product on a scale of 1 to 5. Logistic Regression focuses on predicting categorical target variables (Ray S., 2019).

Logistic Regression has many advantages including its simplicity, speed, and efficiency in training and regularization. It does not require the input characteristics to be scaled. In addition, both data noise and numerous correlations have no effect when using this technique. On the other hand, non-linear issues are not ideal to logistic regression due to its linear decision surface, susceptibility to overfitting, and the necessity of complete knowledge about all independent variables (Ray S., 2019).

D. Linear Regression

A Linear Regression is a kind of supervised learning techniques in which the output value is calculated using the input value and the employed labeled datasets. Continuous variables can be simulated and anticipated using Linear Regression. It also seeks to fit data to a straight hyperplane if the relationship between variables in the data set is linear (Ray S., 2019). It can be computed using equation (2) (Obulesu O et al., 2018):

$$F(x) = mx + b + e \qquad (2)$$

Where F is the dependent variable (x) and x is the independent variable. The line has a slope of m, a y-intercept of b, and an error term of e. Depending on the previous information, a Linear Regression is a simple technique. It is also clear that a linear connection between dependent and independent variables is the most effective. On the other hand, linear regression can only forecast the outcomes. It has a problem with outliers and does not work well with non-linear data. In addition, data must be independent (Obulesu O et al., 2018).

E. Decision Trees

Popular data mining approaches include classifier creation (Kumar R and Verma R., 2012). Classification algorithms can process massive data mining data. The training set and new data can be classified using this strategy. It can also infer categorical class names

(Nikam S., 2015). DTs can perform numerous tasks, including pattern recognition and image processing (G Stein et al., 2005). DT compares numerical attributes to threshold values in various fundamental tests. Tests are ordered and coordinated (I. S. Damanik et al ., 2019). Data mining often classifies using DT (S. S. Gavankar and S. D. Sawarkar., 2017). Every tree has nodes and branches. Each node is a group characteristic that needs to be categorized, and each branch is a candidate value (Mahesh B., 2020). DT is a supervised learning method. A training model that uses learnt decision rules from training data to forecast target variable class or value is the goal (Charbuty B and Abdulazeez A, 2021). Table 1 lists the pros and cons of DT (Y. Zhao and Y. Zhang., 2008 - K. Mittal et al., 2017 - Priyanka and D. Kumar, 2020).

TABLE I. DT BENEFITS AND DRAWBACKS

Advantages	Disadvantages
1- Straightforward. 2- Quickly changed into a set of rules for making things. 3- Can sort both categorical and numerical results but can only create categorical attributes. 4- There are no preconceived ideas about how true the results are.	1-The best way to make decisions can be messed up, which can lead to bad decisions. 2-There are many levels to the decision tree, which makes it hard to understand. 3-As the number of training samples goes up, the calculations of the decision tree may get more complicated.

F. Random Forest

RF uses collective DT strengths in an ensemble classifier to compensate for DT weaknesses (Breiman, L., 2001 - Pal M, 2005 - Cutler D, et al., 2007 - Belgiu M and L. Drăguţ, 2016 - He Y. and T.Warner, 2017). Averaging all tree votes determines each unknown's class. There is no longer concern that planting one tree may not be optimal. Thus, planting a forest should maximize global benefits (Maxwell A. et al., 2018). Bootstrapping builds each tree in the "forest." for resampling. At node splitting, a random subset of characteristics is selected for split variable selection. The majority vote and average determine categorization predictions for regressions (Breiman, 2001 - Yali, A and Donald G, 1997 - Tin K., 1995 - Tin Kam Ho, 1998). RF model tuning uses mtry and ntree. Splitting involves a random number of features (mtry) and several trees (ntree) to determine the output. Trade-offs exist for mtry. Large values make trees comparable but more accurate (Breiman, 2001). Use "Out of Bag" (OOB) samples—unused components—to validate your system. The mean of the trees' OOB-based projections is the result. RFs have a lot of flexibility and a high rate of being right. When the number of trees is considered, it also doesn't fit the data too well. On the other hand, as in DTs, a graphical representation is not possible (Resende P and Drummond A, 2018).

G. Naïve Bayes

The Naive Bayes method is a simple probability-band classifier sometimes known as the Bayes of Idiots, the Bayes of Freedom, or just plain old Bayes. To the extent that the class variable is specified, it is presumed that the presence or absence of a particular class feature is unrelated to the presence or absence of any other class feature (Oladipo I et al., 2022). The Naive Bayes is a simple method of estimating parameters because it does not utilize recursive algorithms. Therefore, a naive Bayes classifier may be of use when working with massive datasets. It also doesn't take a lot of data for training purposes to establish where the boundaries are. The assumption of independence among the variables allows for the estimation of variances rather than the entire matrix of covariance (Oladipo I et al., 2022).

1.2 Unsupervised Learning

Unsupervised learning involves datasets that do not contain any output label information. The objective of these models is to deduce the connection between data and/or to reveal latent variables (Singh G., 2018). High computational cost and multi-collinearity are two issues that can be avoided by limiting the number of features (Dormann F., 2013). During unsupervised learning, the machine makes an educated guess at the real-valued outcome by drawing on its own knowledge and previous encounters. Methods that rely on unsupervised learning include K-Means Clustering, Principal Component Analysis (PCA), and Latent Dirichlet Allocation (LDA) (Singh G., 2018).

A. K-Means Clustering

K-means clustering is a way to learn without being watched. It makes groups or clusters on its own. Data with similar traits are put together in the same cluster. The k-means clustering method has various uses. First, it's easier on a computer than hierarchical clustering with many variables. Second, it yields tighter clusters than global hierarchical clustering and small k. Finally, how simple the clustering results are to use and understand. The algorithm is computationally efficient because its complexity is O (K*n*d) (Ray S., 2019). However, K is uncertain and hard to determine. Global clusters impair performance because different initial partitions produce distinct final clusters. Performance suffers from input data cluster size and number volatility. Correlations between characteristics break the spherical assumption and give related features more weight, making it difficult to distribute features spherically inside each cluster. Outliers can hamper K-Means clustering. Any value of K has a different solution depending on beginning conditions and local ideals. This entails repeating the operation 20–100 times for any K number and selecting the solutions with the lowest J (Ray S., 2019).

B. Principal Component Analysis

PCA is an unsupervised machine learning technique that helps to simplify data. This means the calculations are faster and more reliable (Dey A., 2016). Using PCA, a set of variables can be transformed into a new, orthogonal set of variables known as principal components (PC). Data in two dimensions is reduced to a single dimension. Depending on the scale used, the outcomes of the PCA technique can vary. Therefore, scaling the data set is an essential step in the PCA algorithm (Dhall D and Juneja M., 2020).

C. Latent Dirichlet Allocation

Statistical data mining method LDA can characterize object classes in N-dimensional feature space using a sequence of values. This sequence represents the linear discriminant (McLachlan G., 2005). LDA and PCA are similar in finding the "most relevant" data differences and choosing routes that capitalize on them (I. T. Jolliffe et al., 1986). Labelled class variables distinguish LDA from PCA. How well the class means are divided along that axis compared to the overall class disparities determines its direction. This maximizes the interclass-intraclass dispersion ratio. It explores representations of data with fewer dimensions that cluster examples of the same class and separate instances of other classes (Gow J et al., 2012). Eigenvalues rank the k linear discriminants of the eigenvectors. You can use discriminants to classify new objects or reduce dimensionality (Gow J et al., 2012).

1.3 Reinforced Learning

RL is a goal-oriented technique, in contrast to supervised and unsupervised learning. Humans acquire knowledge through observation of their changing circumstances and responses to those alterations. To accomplish anything useful, RL must rely heavily on a learning agent (controller). Specifically, the agent acts (transmits control signals) to influence the environment's current state and receives unique values representing rewards and punishments. The goal of this agent is to maximize its reward as time progresses. A job provides a detailed description of the environment and elucidates the reward's origin (Coronato A et al., 2012). Instances of RL-based methods include the Q-Learning Algorithm and the Monte-Carlo Tree Search (MCTS).

A. Q-Learning

Using Q-learning (Watkin C & Dayan P.,1992), agents can quickly and easily learn optimal behavior in supervised Markovian environments. It's a low-resource, step-by-step approach to dynamic programming. The system is effective because it steadily improves its evaluations of the quality of certain acts in particular situations. It is possible to classify this technique as a form of Dynamic Programming (DP). Agents can gain insight into the optimal course of action to take in Markovian domains by observing the consequences of their actions without first having to construct detailed maps of those domains (Watkins C., 1989).

Q-learning applications in information theory are being studied. Q-learning and information theory are used in anomaly detection, pattern recognition, natural language processing (NLP), and picture classification (A. Achille and S. Soatto., 2018 - G. Williams et al., 2017 - J.T. Wilkes and C.R. Gallistel., 2017 - Jang B. et al., 2019). As a framework, RL may be used in a voice interaction system to offer a satisfying answer based on the user's speech (Y. An et al., 2017), while DL can predict local rainfall at high resolution (X. Shi et al., 2017). Ant Q-learning emerges in multi-agent settings where agents must agree to calculate a reward for a specific action. Ant Q-learning can get stuck at a local minimum if agents just take the fastest route (Chia-Feng Juang and Chun-Ming Lu., 2009).

B. Monte Carlo Tree Search

MCTS works well for sequential choices. The plan relies on intelligent tree search, which balances exploration and use. Random sampling in simulations helps MCTS track activity data and enhance decision-making with each iteration (Świechowski M., 2021). MCTS makes decisions using tree-based representations of large combinatorial spaces. The nodes of these trees represent problem states or configurations, and their connections show their transitions (Świechowski M., 2021). MCTS is officially used when a Markov Decision Process model can be created. MCTS can be used with POMDP by altering it (Lizotte DJ., 2016). Google DeepMind's AlphaGo (Silver D et al., 2016) is believed to have used MCTS and deep RL.

The MCTS technique is easy to grasp (Browne C et al., 2012). An uneven, systematic tree is built. Each iteration uses a tree policy to find the "root," or most important, node. A balance between discovery and exploitation is key to tree policy. The search tree is updated after a simulation from the node. It needs to update ancestor statistics and add a child node for the specified node's action. Simulation moves follow an unclear default rule. Randomness is the simplest strategy. MCTS minimizes the requirement to evaluate intermediate state values, reducing domain expertise (Browne C et al., 2012).

2. Big Data

The term "Big Data" was introduced in mid-2011 to describe huge, diversified data that is challenging to manage and handle using traditional tools and methods (C. K. Emani et al., 2015). Another definition of "Big Data" is the technology that lets scholars evaluate modern practices' massive data sets (R. L. Villars et al., 2011 - I. A. T. Hashem et al., 2015). BD includes volume, velocity, variety, veracity, and value. Data collected can reach tens or hundreds of petabytes. Velocity refers to the rapid delivery and accumulation of data and the increased need to respond quickly. This definition of "variety" includes semi-structured and unstructured data formats like audio, video, and text. Data like this often needs processing. Data veracity is correctness and reliability. Due to growing data volumes and sources, Big Data can be chaotic in the external world. The ultimate component is value (W. Jiang and J. Luo., 2022). Value is important since it quantifies data's utility (M. Arslan et al., 2017).

Processing data in real-time is a crucial concern in the field of Big Data. The fundamental aspect of BD streaming analytics is in the imperative to examine and react to real-time streaming data, such as traffic statistics, through uninterrupted queries to facilitate

instantaneous analysis during the data flow. The BD stream analysis process commences by ingesting data in the form of an endless tuple. It then proceeds to analyze the data and ultimately generates valuable outputs, typically in the form of an output stream (T. Kolajo et al., 2019).

2.1 Big Traffic Data Sources

This section explores the several data sources that can be employed for traffic prediction. Multiple methodologies exist for categorizing large traffic data. In the past, traffic data was classified into two types: unsupervised and supervised. Supervised data, such as loop detectors and Global Positioning System (GPS) traces, directly provide traffic information. On the other hand, unsupervised data, such as call detail records and cell phone location data indirectly provide traffic information that can be inferred. This evaluation will further classify large traffic data depending on its origins (Sayed S. A. et al., 2023).

Travel Surveys are extensive questionnaires used by municipalities or researchers to collect mobility data. Travel surveys provide comprehensive travel data, reducing the need for traffic forecasts. Travel surveys, albeit useful for real-time traffic forecast, have drawbacks. These include a small sample size, time and location limits, self-reporting inaccuracies, and high data collection expenses (Sayed S. A. et al., 2023). Public travel survey datasets include the Chicago Metropolitan Agency's "My Daily Travel Survey" (T. C. M. A.) and NREL's 2010–2012 "California Household Travel Survey (CHTS)" (T. S. D. Center).

Traffic Sensors data can be collected via radar, acoustics, infrared, and inductive loop detectors. Loop sensors are the most common road sensor technology. They offer many extensively used statistics for traffic prediction. Traffic sensor data can automatically and continually collect massive volumes of data in minutes or seconds. This data has poor spatial coverage and missing data due to insufficient sensors. Additionally, traffic sensor data on traffic flow speed may not be suitable for calculating road segment average journey time. Public traffic survey datasets include the Caltrans Performance Measurement System (PeMS) (Caltrans pems) and the Madrid City Council Open Data Portal's "Traffic Flow Madrid" (M. C. Council).

Electronic Toll Collection (ETC) that are widely utilized on toll highways and toll bridges, as well as in several other applications. The data collected from ETC systems is a thorough and efficient instrument for calculating motorway traffic. An important advantage of ETC systems is the extensive and cost-effective data collection they provide. However, the drawbacks are evident. Toll data can only be gathered in regions where ETC systems are operational (Sayed S. A. et al., 2023). The Amap data for Knowledge Discovery and Data Mining (KDD) CUP 2017 (Tianchi. Kdd cup 2017) is the sole publicly available toll ticket data known to the authors, which is utilized for the prediction of traffic flow.

2.2 Big Data Processing Frameworks

Traditional BD tools in the transportation sector are limited, thus the latest BD technologies have been used to analyze massive amounts of traffic data. Commercial tools have been used to improve spatiotemporal big data support. This section summarizes pre-existing BD processing systems used in traffic prediction jobs or that might be used for these goals. Big Data processing frameworks are either non-spatial or spatial. Non-spatial large data processing frameworks will be divided into Batch, Stream, and Hybrid categories. Figure 2 shows BD processing frameworks for each category (Sayed S. A. et al., 2023).

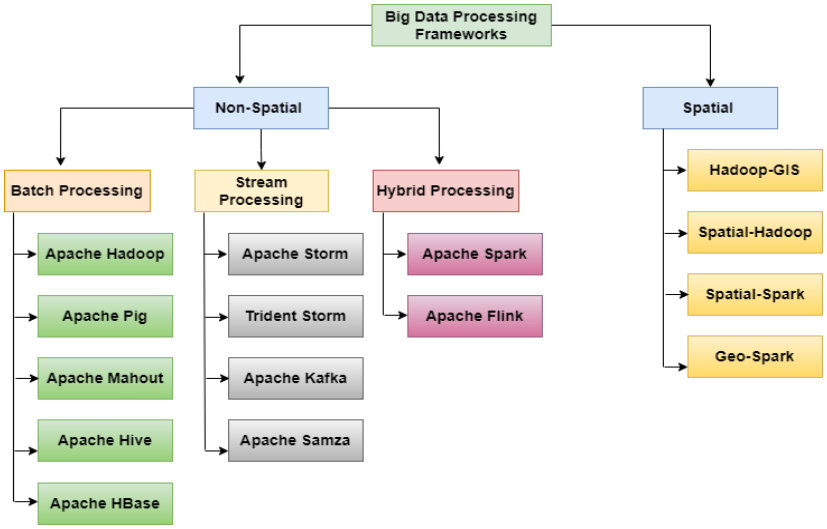

Fig. 2. *BD Processing Frameworks.*

III. SURVEY METHODOLOGY

The articles searched in this review were found in peer-reviewed publications from reputable publishers including Elsevier, Springer, IOP publishing, and IEEE. “Short-term traffic prediction” OR “traffic flow prediction” AND “machine learning” AND “traffic forecasting” OR “traffic speed forecasting” OR “Intelligent Transportation System” AND “Big Data Processing Frameworks”

AND “Big data in traffic flow prediction” are all search phrases that have been used to locate these articles. Prediction, performance evaluation, and the appropriate use of the appropriate techniques are all extensively covered in this paper. The survey's article searched clearly pertains to the topic of using ML methods to forecast traffic patterns. Both primary research articles and reviews were combed through to compile the data shown here.

1. ML Techniques for Traffic Flow Prediction

Predicting future traffic patterns is a crucial part of ITS work (Lippi M. et al., 2013). Many researchers in recent years have examined road traffic data using artificial intelligence, data mining, and statistical methods to forecast future traffic patterns (Aqib M. et al., 2019). As previous studies have shown, no single method is enough to evaluate such massive datasets. Therefore, appropriate technology must be utilized in accordance with the data's structure and volume to extract the most relevant insights (Janković S. et al., 2021).

Regression Model, a machine learning method, predicted traffic patterns in (Deekshetha H. R. et al., 2022). Matplotlib, Pandas, OS, Numpy, Sklearn, Tensorflow, and Keras support the model's regression model. "Traffic prediction" means to anticipate next year's traffic from this year's data. Precision and R2 are shown. Traffic statistics were calculated one hour later. The Kaggle dataset was used for this investigation. Two data sets exist. Some 2015 data includes date, time, number of cars, and number of intersections. The other is 2017 traffic statistics, which precisely match 2016. This study should use deep learning, big data, and other traffic flow prediction methods and examine more forecast accuracy factors.

Suneel Kuamr, 2022 investigated whether it would be possible to solve the problem of traffic regulation by employing an ML technique. The authors simulated traffic signals by employing the Q-learning RL method with an environment that they invented themselves, which they referred to as the Simulation of Urban Mobility (SUMO). The authors’ first set of experiments involved training a model with 300 neurons to determine all the potential parameters, including Cumulative Delay (CD), Average Queue Length (AQL), and Cumulative Negative Reward (CNR). In SUMO, the automobiles that are now in motion can have their delay times monitored, the delay times of individual vehicles can be controlled, and the delay times can be changed.

The main goal of (Navarro-Espinoza A. et al., 2022) was to outline an adaptive traffic control technique. To verify experiment reproducibility, Gradient Boosting Regressor, Linear Regression, MLP Regressor, RF Regressor, and Stochastic Gradient Descendent Regressor were utilized with all other settings set to default and the random state set to zero. In this paper, the authors used Huawei Munich Research Center Road Traffic Prediction Dataset from several traffic sensors. Performance criteria used to evaluate each ML were MAE, MAPE, RMSE, R2, and Explained Variance (EV) Score. When using MLP and Gradient Boosting, R2 and EV were above 0.93, MAE was 10.8, MAPE was 21%, and RMSE was 15.4. However, RF's R2 and EV scores were significantly below 0.93, its MAE 10.88, MAPE 21%, and RMSE 15.5. The RMSE was 15.85, MAE 11.2, and MAPE 24%. EV and linear regression R2 were 0.926. Conclusion: The stochastic gradient has an R2 value of 0.9, EV value of 0.9, MAE value of 12.8, MAPE value of 29%, and RMSE value of 18.

(Upadhyaya S. and Mehrotra D., 2022) integrated four ML models—NB, SVM, DT, and RF—to predict traffic flow lane changes, which is crucial to ITS success. Federal Highway Administration (FHWA) used "Next-Generation Simulation (NGSIM)" to collect a high-fidelity vehicle flow dataset in 2005. Trajectories provided accurate lane designations and vehicle positions. This study uses these data. Data was used to assess model lane-change forecast accuracy. SVM predicted car lane changes better than the other three ML models.

Using the fuzzy logic framework and a time series of urban traffic volumes, the authors of (Qu Z. and Li J., 2022) proposed a type-2 fuzzy logic-based approach to prediction. Interval type-2 fuzzy system predictions were improved with the help of the Back Propagation (BP) technique, which involved updating antecedent coefficients and fuzzy rule outputs. Data on transportation networks were used to assess the method described in this study and compare it to other fuzzy methods. Predictions made with the BP technique and SVM trained using the type-2 fuzzy logic system are more accurate.

In (Steffen T. and Lichtenberg G., 2022), the authors examined historical data-based road traffic prediction methods. This approach uses traffic data convolution polygonal (CP) tensor decomposition. Extraction of daily and weekly features and traffic spatial distribution considerably decreases the quantity of data needed to explain a traffic signal. Deconstructing and moving the key components forward yields traffic data. Starting October 1, 2019, 15-minute data gathering sessions was place on Northern England's M62 until October 28, 2019. Data is shown as vehicles per hour. The four-parameter forecast method outperforms cutting-edge rolling-average prediction algorithms.

An ML-based intelligent traffic monitoring system (ML-ITMS) was developed in (Wang J. et al., 2022) to predict roadside traffic congestion and improve ITS. ML was used to model the current situation, and SVM parameters were adjusted to improve the short-term traffic forecast. The suggested ML-ITMS generated high-capacity, wide-area networks with a single SVM-RF query. The proposed ML-ITMS improved traffic flow and nonparametric process predictions with mathematical models. The proposed ML-ITMS was empirically tested. ML-ITMS applications included automatic parking, LIDAR-based traffic predictions, citywide security monitoring, and healthcare delivery. The results reveal that the proposed ML-ITMS outperforms state-of-the-art approaches in traffic tracking (up to 98.6% accuracy) and traffic flow prediction.

A geophysical-informed search approach for an extreme learning machine (ELM) was proposed in (Cui Z. et al. 2022). Short-term traffic forecasts are being considered as a potential answer. ELM avoids BP's tedious process by zeroing in on the best possible solution. Taking everything into account, the suggested search approach aims to find the best possible settings for ELMs. Many recent models have been used to assess how well the proposed search approach predicts on four standard data sets. The real traffic on the A1, A2, A4, and A8 freeways that make up the Amsterdam Ring Road was used to create the four reference datasets. On these datasets, the GSA-ELM model obtains MAPEs of 11.69%, 10.25%, 11.72%, and 12.05%, while the RMSEs fall in the 288.09%-163.24% range.

Supervised ML was investigated as a Big Data analytics technique for traffic volume forecasting in (Janković S. et al., 2021). Two case studies were used to determine this. Both sets of studies employed training and testing data gathered from prediction models fed traffic counts collected by selected automatic traffic counters deployed on roadways in the Republic of Serbia between 2011 and 2018. In the first study, models based on DT got the best results. In the second study, models based on Lazy IBk, RF, Random Committee, and Random Tree algorithms got some of the best results. In the first case study, the model based on the M5P algorithm demonstrated the highest performance, whereas the Lazy IBk method produced the best results in the second study.

To organize traffic, the authors of (Li J. et al., 2021) developed an ML method for estimating journey times. They tested the Gaussian Process Regressor (GPR) for this. After determining the average travel time for a route, "clustering" shows that different "kinds of days" have varied daily travel time profiles. Then, numerous regression parameters estimated traffic volume from journey times. The authors examined two cases in this study. The regression component for each day's profile was trained using "multi-model" variance. This "Single Model" variation trained one Regressor without considering the day profile. The unique route network traffic flow prediction and reestablishment strategy uses ML and automobile data. This study illuminates two main difficulties. First, training with non-dispersed algorithms might complicate and lengthen assessment sequences. Second, Every ML-based approach has a second issue regarding data quality relevance.

In (Knapińska A. et al., 2021), the authors examined online multiple-time series prediction for traffic of different frame sizes. The model predicted traffic using previous data. The authors began by explaining how they collected and analyzed real-world network traffic data to find periodicity and relationships across traffic types. They also used linear regression, KNN, and RF to predict network data under different models and input features. Linear regression has 50% lower RMSPE than KNN and 15% lower than RF. The authors also examined how different models and input features might affect the outcome, finding the best speed-accuracy balance.

In (Zeroual A. et al., 2021), traffic flow predictions were made using a kernel-based learning model called support vector regression (SVR). To evaluate the SVR model's ability to predict traffic densities, the authors looked into a variety of kernels. Data collected from California motorways was used to assess the SVR's effectiveness. The authors also provided a comparison between linear regression, linear SVR, quadratic SVR, and gaussian SVR. Results from this study supported the use of the Gaussian SVR for predicting traffic flows. These findings demonstrate that the Gaussian kernel SVR model is superior to the alternatives.

In (Ramchandra N. R. and Rajabhushanam C., 2021), many well-established ML methods, such as Deep Autoencoder (DAN), Deep Belief Network (DBN), and RF, were applied to an online data set to forecast traffic flows. It was possible to foretell the zone's traffic volume by using the four most important characteristics of time of day, season, temperature, and location. Accuracy, precision, RMSE, and MSE values were used to assess the proposed system's efficacy. The RF method is the most effective of the three with an accuracy of 92.6%, MSE of 346.35, and RMSE of 18.61.

Global and cluster-based methods for learning vehicle speed prediction in large, dynamic sensor networks were given in (Magalhaes R. P. et al., 2021). The authors evaluate and contrast prediction models trained with the three techniques to answer three experimental questions. The authors employed a large real-world sensor dataset and cutting-edge ML methods to train the model. We tested multivariate linear regression (MLR), radial basis functions (RF), and gradient boosting regression trees (GBRT). Over the dataset's 12 months, 130 sensors were added to the system. Developing global models for dynamic sensor network difficulties has shown success. Researchers received the real-world dataset necessary to validate their findings. Fortaleza, Brazil, had 272 road traffic sensors that collected 1.3 billion data points in 2014. This was the study's dataset. Historical Average (HA), a baseline method for estimating average speed in a time interval by averaging all training example speeds, was also studied.

Trend-based Online Network Traffic Analysis (TONTA), a new pattern recognition method for ad-hoc IoT networks, was suggested in (Shahraki A. et al., 2021) to monitor network performance. A statistically light method called Trend Change Detection (TCD) was used online in the proposed method. To analyze the traffic on the IoT network, TONTA looks for major trends and notices sudden or gradual changes in time-series datasets. As an offline benchmark TCD technique, Relative unconstrained Least-Squares Importance Fitting (RuLSIF) was used to compare how well TONTA worked with how well it worked. The results showed that TONTA finds about 60% fewer false positive alarms than RuLSIF.

A hybrid model using ELM and ensemble-based approaches was constructed to forecast future hourly traffic on a road stretch in Tangiers, northern Morocco (Jiber M. et al., 2020). SLFN and other fast ML algorithms were employed to develop their strategy. Moroccan Center for Road Studies and Research annualized traffic estimates from 2013 to 2017 were used to complete the work. Consider all factors that affect road traffic flow to gain a complete picture. Conditions like weather and highway layout are examples. The suggested model was compared to MLP, SVR, and ARIMA standard algorithms. Simulations revealed good precision and stability.

Four ML models—RF, SVR, MLP, and MLR—were tested for predictive power using road network data from Thessaloniki, Greece (Bratsas C. et al., 2020). A vast collection of stationary and mobile traffic data covered the metropolis in space and time. The authors summarized 15-minute periods of numbers. They standardized the number of cars entering each route, its minimum and highest speeds, standard deviations, means, skewnesses, and kurtosis, and used these statistics to train ML algorithms that forecast average speeds. The optimal model parameters were found using 10-fold cross-validation. Splitting the training dataset into 10 similar datasets allowed 10-fold cross-validation. The requested mean speed estimates were made by training the RF, SVR, and MLP on 9 datasets. The SVR model works best in stable, low-change contexts, whereas the MLP model is more flexible and has near-zero errors in high-change environments.

The authors of (Meena G. et al., 2020) devised an accurate and timely traffic flow prediction algorithm. Combining ML, genetic, and soft computing methods simplified transportation system data analysis. Traffic sign recognition, a key stage in autonomous vehicle training, used image processing. The paper's goals were met utilizing Android Studio, Java, Garmin, PHP, XML, Python, and sklearn. The authors constructed and analyzed many ML methods to improve productivity and precision. The DT algorithm predicted target variable values (88% accuracy, 108.4 sec). They used SVM (Accuracy: 88%, Time: 94.1 sec) to locate the most extreme data for a precise forecast. Finally, the RF method (Accuracy: 91%, Time: 110.1 sec) predicted traffic congestion.

In (Zheng L. et al., 2020), the authors proposed a gradient-boosted regression tree (DSTO-GBRT) based on dynamic spatial-temporal characteristics for short-term traffic flow prediction using Electronic Registration Identification (ERI) big data, a new vehicle identification technology. First, the DSTO-GBRT architecture was created. The spatial-temporal relationships between the current forecast point and upstream correlative locations were examined using Pearson's correlation coefficient. PCA minimized linear correlations among features to refine the initial training data and generate high-quality data. The authors tested the model using Chongqing ERI data from March 1 to March 31, 2016. Test data was used on March 25, 2016, and training data was used elsewhere. Compared to ST-GBRT, ARIMA, DSTO-BPNN, and DSTO-SVM, DSTO-GBRT provided a rapid and adaptable prediction, especially during rush hour when traffic conditions change quickly. The proposed DSTO-GBRT method surpassed DSO-GBRT and DTO-GBRT in accuracy.

In (Kamble S. J. and Kounte M. R., 2020), multiple characteristics, including hard delay limitations and the speed available across a GPS vehicle's route were used to identify traffic congestion using the ML technique. To estimate traffic speeds, they have implemented the Gaussian Regressor process in ML, which makes use of three datasets: the training set, the prediction set, and the road sector data frame for historical route networks data. The authors have used the data set to determine three distinct periods during which to monitor traffic congestion and have analyzed the average speed of cars on the road sector during each of these periods.

A revolutionary architecture predicted city traffic flow, according to (Inzunza M. C. et al., 2020). ML, Computer Vision (CV), DL, and neural networks were considered for the answer. The Machine Algorithm System (MASY) analyzes traffic patterns, the Neuronal Artificial System (NASY) classifies traffic, the Web user application (WeUsAP) displays results and processes user data entry, and the Car Counting Wizard (CCW) captures video using computer vision to generate a statistical analysis of vehicle counts. The approximation was accurate, the model improved over time, and the forecast model matched CCW's automotive output.

The authors of (Weerasekera R. et al., 2020) tested ANN, RF, and SVR algorithms for simulating traffic flow at various data resolutions and responding to unexpected traffic accidents. To focus on spatiotemporal attributes that most affect model accuracy, several feature selection methods have been examined. These tests showed that multivariate spatiotemporal ML models with aggregated data are not always effective. Models learned with high-resolution 30-second input data outperformed baseline ARIMA models by 8%. An added benefit was that Recursive Feature Elimination-based feature selection beat linear correlation-based models. The authors used Auckland roadway traffic and occupancy information to derive conclusions.

The authors of (Zahid M. et al., 2020) used state-of-the-art models and hyperparameter optimization to accurately simulate short-term future traffic condition prediction using ML classifiers such local deep SVM (LD-SVM), decision jungles (DJ), MLP, and CN2 rule induction. Level of service (LOS) horizons and simple if-then rules are also used to assess traffic conditions over time. They found that random hyperparameter tweaking produced the best results. Decision jungle and LD-SVM improved prediction accuracy by 0.982 and 0.975, respectively, for a roughly 95% improvement. The experiments indicated that DJs were more trustworthy and productive than competitors. VISSIM was used to simulate a basic section of Beijing's Second Ring Road expressway and calculate short-term traffic status as a function of level-of-service (LOS). A roadway length's real density flow was calculated using 15, 10, and 5-minute prediction times.

A hybrid predicting model that combines decomposition and prediction was introduced in (Wang Z. et al., 2020) to improve highway traffic flow estimation. The California Department of Transportation's Performance Measurement System detector VDS-1209092 on Irvine's I405-N motorway provided the data. They trained and adapted prediction models using 2016 data points from May 1 to May 7, 2019, including speed, occupancy rate, and traffic flow. Every data set was split in two. This model was trained using data from May 1–5. The model's settings were adjusted on May 6 and 7, 2019. After training the model, the day's traffic volume was predicted and compared to May 8, 2019. The full ensemble empirical mode decomposition with adaptive noise (CEEMDAN) approach adaptively decomposed complicated, nonlinear highway traffic flow data. Improved weighted permutation entropy was used to reconstruct new elements. The least-squares SVM (LSSVM) prediction model they created for each reconstruction segment

included the subsequence predictions, making it more accurate. The authors' experiments showed that the model helps anticipate traffic flow and analyze trends, helping transportation officials make better decisions.

Intelligent Internet of Things (IIoT) networks are dynamic due to their ever-changing topology and large range of services, making traffic forecasting difficult. These observations encouraged the authors in (Nie L. et al., 2020) to propose a reinforcement learning mechanism. They predicted network traffic using a Markov decision process model and refined it with Monte-Carlo Q-learning. They also devised a residual-based dictionary learning technique to simplify Monte-Carlo Q-learning and made the process real-time. The authors tested the RL-based network traffic prediction algorithm on a 12-node testbed. The constructed wireless network supports video, telephony, and other services. The generic urban path loss model sets OSPF weights, and the open shortest path first (OSPF) algorithm creates the testbed's topology. Their proposed technique was tested using real network traffic.

Clustering and series models were combined in (Aldhyani T. H. et al., 2020) to improve network traffic prediction. They improved our time series using fuzzy c-means clustering granules. The suggested model was tested using real network data (4G Cell Traffic from Kaggle and Measurement and Analysis on the Wide Internet) from diverse network backbones. The proposed research uses clustering to handle ambiguity in entire network data, which is better than time series models. For even greater model accuracy, the authors recommend preprocessing using a weighted exponential smoothing model. The authors claimed AI could predict network user behavior. AI is preprocessing time series models to improve them. In the Adaptive Neuro-Fuzzy Inference System (ANFIS), non-crisp Fuzzy-C-Means (FCM) clustering and weight exponential technique produce better time series models. To predict network traffic, ANFIS time series model was constructed. Two snapshots of network traffic were used to evaluate the suggested time series models. ANFIS model empirical tests utilizing cellular traffic data showed a correlation indicator R of 96.78%. The suggested model outperforms competing time series models.

EL was used to update the distributional representation in (Xiao J. et al., 2019), a new stepwise regression method in an ideal drift environment based on Learn++ for SVR. The author first converted the regression problem into a binary classification problem to estimate future foot traffic. Second, the R2C technique improved classification-style ensemble learning loss functions. Next, represent hyper-resolution improvement as incremental regression function learning. Since motion volume is spatially dependent, the R2C architecture's motion volume prediction is flawed. This study used data from the Caltrans PeMS, which receives traffic data from over 15,000 detectors every 30 seconds. California's motorways had detectors everywhere.

A new ITS infrastructure approach was suggested in (Rajkumar S. C., 2019). This approach used a magnetic sensor to count and classify automobiles in a traffic pattern. The cluster then delivers data to a cloud server using MapReduce and local proximity services. An intelligent cloud server agent using the Markov Decision Process (MDP) reinforcement learning algorithm predicted the best route for registered users. This study employed a real-time dataset from 1 Nov to 15 Nov 2018. Applied successfully, this approach has a 98.36% success rate.

In (Tu M. et al., 2019), the authors introduced a new traffic prediction method using least squares SVM and K-means clustering algorithms for data-driven traffic management and decision-making. Historical traffic statistics came from Nanchang's Donghu District. The LSSVM algorithm has a prediction error rate of 25.33 %, while K-means clustering had 20%. Due to traffic flow data anomalies, transportation system complexity, and data incompleteness, there are still many defects. (1) Traffic flow data volume and connection. (2) Flow, vehicle speed, and occupancy aspects of traffic flow data are interdependent. (3) The model ignores abrupt events like traffic accidents and extreme weather. The specific qualities of highways, trails, and national roads, which should be considered, were not researched.

Traffic-noise discomfort was predicted using ML models (Bravo-Moncayo L. et al., 2019) that incorporated noise perception, exposure, and demographics. ANN, SVM, and MLR were used to create traffic noise nuisance models and compare their error rates. City building data was obtained from the Municipality database registration for this investigation. From 2010 to 2016, the Municipal Mobility Secretariat conducted one-week automatic car counts on 523 streets on various roadways. The case study area's traffic noise exposure was mapped and estimated. Since an individual's perception of noise and expected exposure to noise affect traffic-noise discomfort, traditional statistical models cannot provide trustworthy forecasts. Implementing an ML technique enhanced accuracy and R2 values. The ANN model predicted traffic-noise annoyance better than the MRL and SVM by 42% and 35%, respectively, in training subsets. Test dataset subsets had 24% and 19% less error. ANN had 3.8- and 2.3-fold R2 gains in training datasets and 1.7-fold increases in testing datasets compared to MLR and SVM models. In this study, MatLab and R were used to tackle classification and regression issues using MLA.

In (Li Y. and Jiang W., 2019), big data short-term traffic flow prediction accuracy and timeliness were examined. The Caltrans PeMS contributed this study's data. Traffic data is collected in real time from over 39,000 detectors. This report uses data from Pasadena Road Monitoring Point 771668 from January 1, 2010 to January 1, 2019. This study's data samples use a 5-minute sale detection time. Cluster analysis and Regression Forest were used to develop a Spark prediction model. It used K-means cluster analysis to assess traffic volume and climatic conditions over time and space. The parallel Regression Forest method was used to predict and train massive data sets on a spark distributed computing cluster. Testing showed that the integrated forecast model for comparable data sets runs faster on a Spark cluster. The benefit grows with data volumes. When employed alone, it outperforms SVM and regression forest in forecast accuracy.

In (Fang C. et al., 2019), the authors found that identifying road traffic statuses significantly affected short-term traffic flow forecast accuracy. This study uses data from six RTMS on Beijing's second ring road during December 1–5, 2003. The experiment's RTMS detector is located on Beijing's four-lane Second Ring Road. A workday is the timeframe. The authors anticipate traffic flows using ARMA and Kalman filters in their study. The suggested method splits measured traffic data for each road traffic condition in half and compares the halves using expected outcome indices. Finally, many traffic forecasts and flow advice were given. This study found that state partitioning drastically reduces ARMA and Kalman filter short-term traffic flow forecast accuracy. A plus is that an ARMA model can be adjusted to increase forecast accuracy when road traffic tidal characteristics are clear and lane conditions are the same. Predictions and results may vary depending on state partitioning approach.

In (H. Mehdi et al., 2019), cloud traffic was predicted using a method called fuzzy autoregressive integrated moving average (FARIMA), which combines ARIMA with fuzzy regression. In this study, the authors used a portion of the Wideadjp dataset for simulation. From November 1 to December 30, 2015, Wideadjp received 228 GB requests. Better prediction accuracy can be attained with the use of fuzzy ARIMA (SOFA) models. RMSE and coefficient of determination comparisons show that SOFA is the more accurate model for traffic prediction, with values of approximately 5.4 and 0.009, respectively.

Backpropagation bidirectional ELM (BP-BELM) models are introduced in (Zou W. and Xia Y., 2019) as a unique prediction model. To fine-tune back propagation settings, prior experience is not required. In this study, the authors used eight UCI datasets (AutoMPG, Automobile, DrivFace, Fertility, NoisyOffice, Servo, UJIIndoorLoc, and wiki4HE). In the end, the simulations, and comparisons show that BP-BELM is more accurate at predicting traffic flow than other methods, such as back propagation neural networks, radial basis functions, SVM, and others that have been developed incrementally.

The authors of (Yang H. et al., 2019) employed the powerful and systematic Taguchi approach to optimize the proposed exponential smoothing and ELM forecasting model and find the optimal configuration. The new model was tested on data collected from UK motorways and freeways, and the results were compared to those of three other forecasting models. With an accuracy of about 91% and 88% during peak and off-peak traffic times on the freeway and highway, respectively, the results showed that the Taguchi technique is a great way to design a forecasting model and that the proposed model with its optimized configuration does a better job of predicting traffic flow. To summarize all previous related works, Table 1 compares them in terms of methodology, data set, adopted techniques, and results.

TABLE II. SUMMARY OF RELATED WORKS.

No.	Methodology	Dataset	Techniques	Results
(Deekshetha H. R. et al., 2022)	ML algorithm to forecast next year's traffic using data from the previous year.	Two Kaggle datasets.	Regression Model	N/A.
(Suneel Kuamr, 2022)	Traffic signals simulation with an environment called SUMO.	N/A.	Q-learning RL method.	SUMO monitors, controls, and changes delay times
(Navarro-Espinoza A. et al., 2022)	A strategy for adaptive traffic control.	Huawei Munich Research Center Dataset.	Linear Regression, MLP Regressor, Gradient Boosting Regressor, RF Regressor, and Stochastic Gradient Descendent Regressor.	R^2: of 0.9, EV: 0.9, MAE: 12.8, MAPE: 29%, and RMSE: 18.
(Upadhyaya S. and Mehrotra D., 2022)	Predict lane changes in traffic flows.	High-fidelity vehicle flow dataset in 2005.	SVM, NB, RF, and DT.	SVM has the highest accuracy.
(Qu Z. and Li J., 2022)	Time series analysis of urban traffic volumes prediction	Data on transportation networks.	Interval type-2 fuzzy logic, BP, and SVM.	Predictions with type-2 fuzzy logic system are more accurate.
(Steffen T. and Lichtenberg G., 2022)	Predicting traffic along a certain road based on past data.	Northern England's M62 dataset.	CP tensor decomposition of traffic data.	The proposed forecast method significantly beats the competition.
(Wang J. et al., 2022)	ML-ITMS to predict congestion at roadside sensors.	N/A	SVM and RF.	Accuracy 98.6%
(Cui Z.et al., 2022)	Short-term traffic predictions based on ELM.	Real traffic data on the Amsterdam Ring Road	ELM.	**MAPEs:** of 11.69%, 10.25%, 11.72%, and 12.05%, **RMSEs**: 288.09% to 163.24%
(Janković S. et al., 2021)	Supervised ML traffic volume prediction.	Counts of vehicles on the roads in the Republic of Serbia.	DT, Lazy IBk, RF, Random Committee, and Random Tree algorithms.	- M5P and Lazy IBk have the best results
(Li J. et al., 2021)	An ML technique for predicting journey times.	N/A.	GPR.	N/A.
(Knapińska A.et al., 2021)	An online multiple-time series traffic prediction.	N/A.	Linear Regression, KNN, and RF.	Linear Regression has highest accuracy.
(Zeroual A. et al., 2021)	SVR model to forecast traffic flow.	Caltrans PeMs.	Gaussian SVR.	Gaussian SVR has the best performance.
(Ramchandra N. R. and	Forecast traffic flows using ML methods.	An online dataset.	DAN, DBN, RF.	RF: 92.6% accuracy.

Rajabhushanam C., 2021)				
(Magalhaes R. P. et al., 2021)	Global and cluster-based learning for vehicle speed forecasting.	A sizable sensor dataset spans 12 months.	MLR, RF, GBRT.	HA was investigated
(Shahraki A. et al., 2021)	TONTA to monitor network performance using TCD.	Traffic data on the IoT network.	TCD, RuLSIF.	TONTA: 60% fewer false positive than RuLSIF.
(Jiber M. et al., 2020)	Hourly traffic forecasting.	Annualized traffic numbers from 2013 to 2017.	SLFN	The proposed model offers higher performance.
(Bratsas C. et al., 2020)	Evaluate the predictive skills of some ML models.	Data from the road network in Thessaloniki, Greece.	RF, SVR, MLP, and MLR.	SVR and MLP have the best performance.
(Meena G. et al., 2020)	ML, genetic, and soft computing approaches for predicting traffic flow information.	Traffic signs recognized using Image Processing techniques.	DT, SVM, RF	**RF**: Accuracy 91%, Time 110.1 sec.
(Zheng L. et al., 2020)	DSTO-GBRT to predict traffic flows	Real-world big ERI data from March 1 to March 31 of 2016	GBRT, Pearson's correlation coefficient, and PCA.	DSTO-GBRT has the best performance
(Kamble, S. J. and Kounte M. R., 2020)	ML technique to identify traffic congestion.	Road sector data frame for historical route networks data.	Gaussian Regressor.	The authors have analyzed the average speed of cars on the road sector.
(Inzunza M. C. H. et al., 2020)	A novel architecture was built to anticipate traffic flow.	N/A.	ML, CV, DL, and neural networks.	The model has heigh accuracy.
(Weerasekera R. et al., 2020)	ML techniques to model traffic flow.	Traffic records from a roadway in Auckland (New Zealand).	ANN, RF, and SVR.	The proposed models have heigh accuracy
(Zahid M. et al., 2020)	Cutting-edge models and hyperparameter optimization to predict traffic conditions.	A simple part of the Second Ring Road freeway in Beijing, China.	LD-SVM, DJ, MLP, and CN2 rule induction.	Decision jungle and LD-SVM have the best performance.
(Wang Z. et al., 2020)	Hybrid predicting model for traffic prediction.	Caltrans PeMS Dataset.	LSSVM.	The model aids in traffic flow forecasting and trend analysis.
(Nie L. et al., 2020)	RL-based network traffic prediction method.	A testbed with 12 nodes to test the proposed model	RL, Q-Learinig.	An analysis of actual network traffic was used to test the model.
(Aldhyani T. H. et al., 2020)	ANFIS time series model was created for network traffic prediction.	4G Cell Traffic data from Kaggle.	Fuzzy-C-Means clustering.	correlation indicator R: 96.78%.
(Xiao J.et al., 2019)	A new framework based on stepwise regression for traffic prediction.	Caltrans PeMS.	Learn++ for SVR, R2C method,	R2C architecture's low accuracy
(Rajkumar S. C., 2019)	A new ITS infrastructure to determine both the vehicles count and their types in traffic pattern.	A real-time dataset from 1 Nov to 15 Nov 2018.	RL.	**RL**: 98.36% accuracy.
(Tu M. et al., 2019)	A new method for traffic prediction.	Historical traffic data from Nanchang's Donghu District.	LSSVM and K-means clustering	**LSSVM**: error rate 25.33 %, **K-Means:** error rate 20 %.
(Bravo-Moncayo et al., 2019)	Traffic-noise discomfort was predicted using ML models.	Information about buildings in cities from the Municipality database register.	ANN, SVM, and MLR.	The ANN model has the highest accuracy.
(Li Y. et al., 2019)	traffic flow predictions using Spark cluster.	Caltrans PeMS.	K-means clustering.	the proposed model has a high accuracy
(Fang C. et al., 2019)	The precision of short-term traffic flow forecasts was investigated.	Information gathered by six RTMS on Beijing's second ring road.	ARMA, Kalman filter technique	prediction accuracy of the ARMA and Kalman filter models affected by state partitioning.
(H. Mehdi et al., 2019)	Cloud traffic was predicted using FARIMA.	A portion of the Wideadjp dataset for simulation.	ARIMA, fuzzy regression.	**RMSE and coefficient of determination**: 5.4 and 0.009, respectively.
(Zou W. and Xia Y., 2019)	ELM (BP-BELM) models were developed as a unique prediction model.	Eight UCI datasets	BP-BELM.	BP-BELM has a high accuracy.
(Yang H. et al., 2019)	ELM forecasting model	Data from UK motorways and freeways.	ELM.	ELM: 91% Accuracy.

IV. CHALLENGES AND OPPORTUNITIES

Traffic bottlenecks can be dangerous, especially in heavily populated places, thus planning and predicting traffic is crucial. Reliable and effective road traffic prediction methods are needed. This study addresses the absence of computer-friendly traffic flow forecasting methods and algorithms. High-quality training data is scarce. Network models were trained using incomplete data from matched city traffic flow statistics. If they are true, ML can't estimate traffic flow as well. ML's underuse of real-time spatio-temporal correlations causes the gap. The complicated linkages between road segments and congestion patterns explain this. Lack of processing power and centralized storage complicates traffic forecasts. This needs further study.

The current study is limited to the processes and algorithms in the examined literature. It's possible this investigation overlooked some techniques. The literature study covers powerful DL methods like CNN and LSTM, which should be studied further. This is now

achievable since traffic data from different cities may be utilized to train models on more universal data patterns. Thus, ML and DL algorithms will improve traffic projections in smaller cities. Communicating with the city's urban administration to deliver a mountain of relevant big data will be the researchers' biggest challenge. Legal constraints on sharing traffic data to municipal governments can provide challenges. Traffic-monitoring sensors fed into machine-learning algorithms for better decision-making may increase linked, high-risk IoT environments. Cybersecurity threats should be anticipated in ITS cities. This raises several research questions.

There have been many successful Big BD traffic prediction applications, however there are still certain challenges. The data density varies widely between modes of transportation, and data shortage, excessive missing data, distortion, and deficiency persist. Data quality, privacy, and policies have been understudied. This section discusses possible solutions. Crowdsourced data has poor quality, noise issues, and privacy concerns. To circumvent these issues, sparse BA has been used to predict traffic conditions using under-sampled data (C. N. Babu et al., 2019).

V. CONCLUSION

The objective of this paper was to provide a thorough analysis of the prominent ML approaches employed in traffic forecasting, while also examining the latest advancements in big data in traffic prediction. The focus was on the processes of traffic prediction and the suitability and effectiveness of ML and BD as viable options. Additionally, it provided an examination of the many sources of big data that can be employed in traffic forecasting activities. Furthermore, it addressed the challenges linked to the utilization of BD and ML in traffic prediction. As a result of this review, a total of 36 papers were selected and thoroughly investigated after a rigorous selection procedure.

REFERENCES

[1] A. Achille and S. Soatto, "Information dropout: Learning optimal representations through noisy computation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[2] Aldhyani, T. H., Alrasheedi, M., Alqarni, A. A., Alzahrani, M. Y., & Bamhdi, A. M. (2020). Intelligent hybrid model to enhance time series models for predicting network traffic. IEEE Access, 8, 130431-130451.

[3] An, S., Lee, B.-H., & Shin, D.-R. (2011). A Survey of Intelligent Transportation Systems. 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks.

[4] Aqib, M.; Mehmood, R.; Alzahrani, A.; Katib, I.; Albeshri, A.; Altowaijri, S.M. 2019. Smarter Traffic Prediction Using Big Data, In-Memory Computing, Deep Learning and GPUs, Sensors 19: 2206.

[5] Belgiu, M., and L. Drăguţ. 2016. "Random Forest in Remote Sensing: A Review of Applications and Future Directions." ISPRS Journal of Photogrammetry and Remote Sensing 114: 24–31.

[6] Bengio, Y., Learning deep architectures for AI. 2009: Now Publishers Inc.

[7] Bratsas, C., Koupidis, K., Salanova, J. M., Giannakopoulos, K., Kaloudis, A., & Aifadopoulou, G. (2020). A comparison of machine learning methods for the prediction of traffic speed in urban places. Sustainability, 12(1), 142.

[8] Bravo-Moncayo, L., Lucio-Naranjo, J., Chávez, M., Pavón-García, I., & Garzón, C. (2019). A machine learning approach for traffic-noise annoyance assessment. Applied Acoustics, 156, 262-270.

[9] Breiman, L. 2001. "Random Forests." Machine Learning 54 (1): 5–32.

[10] Breiman. 2001. Random forests. Machine Learning 45, 1 (Oct. 2001), 5–32.

[11] Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., ... & Colton, S. (2012). A survey of monte carlo tree search methods. IEEE Transactions on Computational Intelligence and AI in games, 4(1), 1-43.

[12] C. K. Emani, N. Cullot, and C. Nicolle, "Understandable big data: a survey," Computer science review, vol. 17, pp. 70–81, 2015.

[13] C. N. Babu, P. Sure, and C. M. Bhuma, "Sparse bayesian learning assisted approaches for road network traffic state estimation," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 3, pp. 1733–1741, 2020.

[14] Caltrans. Caltrans performance measurement system (pems). [Online]. Available: (http://pems.dot.ca.gov/).

[15] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.

[16] Chen, C., Li, K., Teo, S. G., Zou, X., Li, K., & Zeng, Z. (2020). Citywide traffic flow prediction based on multiple gated Spatio-temporal convolutional neural networks. ACM Transactions on Knowledge Discovery from Data (TKDD), 14(4), 1–23.

[17] Chen, K., Chen, F., Lai, B., Jin, Z., Liu, Y., Li, K., Wei, L., Wang, P., Tang, Y., Huang, J., Hua, X. (2020). Dynamic Spatio-temporal graph-based CNNs for traffic flow prediction. IEEE Access, 8, 185136–185145.

[18] Chia-Feng Juang and Chun-Ming Lu. 2009. Ant colony optimization incorporated with fuzzy Q-learning for reinforcement fuzzy control. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 39, 3 (2009), 597–608.

[19] Chowdary, G. J. (2021). Machine Learning and Deep Learning Methods for Building Intelligent Systems in Medicine and Drug Discovery: A Comprehensive Survey. arXiv preprint arXiv:2107.14037.

[20] Coronato, A., Naeem, M., De Pietro, G., & Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. Artificial Intelligence in Medicine, 109, 101964.

[21] Cui, Z., Huang, B., Dou, H., Tan, G., Zheng, S., & Zhou, T. (2022). GSA-ELM: A hybrid learning model for short-term traffic flow forecasting. IET Intelligent Transport Systems, 16(1), 41-52.

[22] Cutler, D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. "Random Forests for Classification in Ecology." Ecology 88 (11): 2783–2792.

[23] Deekshetha, H. R., Shreyas Madhav, A. V., & Tyagi, A. K. (2022). Traffic Prediction Using Machine Learning. In Evolutionary Computing and Mobile Sustainable Networks (pp. 969-983). Springer, Singapore.

[24] Dey, A.: Machine learning algorithms: a review. Int. J. Comput. Sci. Inf. Technol. 7(3), 1174–1179 (2016).

[25] Dhall, D., Kaur, R., & Juneja, M. (2020). Machine learning: a review of the algorithms and its applications. Proceedings of ICRIC 2019, 47-63.

[26] Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1), 27-46.

[27] Fang, C., Gao, D., Xue, Y., & Xiong, Z. (2019, July). Research on short-term traffic flow prediction method based on real-time traffic status. In IOP Conference Series: Materials Science and Engineering (Vol. 569, No. 5, p. 052062). IOP Publishing.

[28] G. Stein, B. Chen, A. S. Wu, and K. A. Hua, "Decision tree classifier for network intrusion detection with GA-based feature selection," in Proceedings of the 43rd annual Southeast regional conference-Volume 2, 2005, pp. 136–141.

[29] G. Williams, N. Wagener, B. Goldfain, P. Drews, J.M. Rehg, B. Boots, and E.A. Theodorou, "Information-theoretic mpc for model-based reinforcement learning," In Robotics and Automation (ICRA), IEEE International Conference on, pp. 1714–1721, 2017.

[30] Gow, J., Baumgarten, R., Cairns, P., Colton, S., & Miller, P. (2012). Unsupervised modeling of player style with LDA. IEEE Transactions on Computational Intelligence and AI in Games, 4(3), 152-166.

[31] H. Mehdi, et al., Cloud traffic prediction based on fuzzy arima model with low dependence on historical data, Trans. Emerg. Telecommun. Technol. (2019) e3731.

[32] He, Y., E. Lee, and T. A. Warner. 2017. "A Time Series of Annual Land Use and Land Cover Maps of China from 1982 to 2013 Generated Using AVHRR GIMMS NDVI3g Data." Remote Sensing of Environment 199: 201–217.

[33] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," Information systems, vol. 47, pp. 98–115, 2015.

[34] I. S. Damanik, A. P. Windarto, A. Wanto, S. R. Andani, and W. Saputra, "Decision Tree Optimization in C4. 5 Algorithm Using Genetic Algorithm," in Journal of Physics: Conference Series, 2019, vol. 1255, no. 1, p. 012012.

[35] I. T. Jolliffe, Principal Component Analysis. New York: SpringerVerlag, 1986.

[36] Inzunza, M. C. H., Robles, L. H., Carlos-Mancilla, M. A., & López-Neri, E. (2020). Traffic Prediction Architecture based on Machine Learning Approach for Smart Cities. Res. Comput. Sci., 149(11), 23-33.

[37] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: a survey," IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1624–1639, 2011.

[38] J.T. Wilkes and C.R. Gallistel, "Information theory, memory, prediction, and timing in associative learning," Computational Models of Brain and Behavior, pp. 481–492, 2017.

[39] Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. IEEE Access, 7, 133653-133667.

[40] Janković, S., Uzelac, A., Zdravković, S., Mladenović, D., Mladenović, S., & Andrijanić, I. (2021). TRAFFIC VOLUMES PREDICTION USING BIG DATA ANALYTICS METHODS. International Journal for Traffic & Transport Engineering, 11(2).

[41] Janković, S., Uzelac, A., Zdravković, S., Mladenović, D., Mladenović, S., & Andrijanić, I. (2021). TRAFFIC VOLUMES PREDICTION USING BIG DATA ANALYTICS METHODS. International Journal for Traffic & Transport Engineering, 11(2).

[42] Jiber, M., Mbarek, A., Yahyaouy, A., Sabri, M. A., & Boumhidi, J. (2020). Road traffic prediction model using extreme learning machine: the case study of Tangier, Morocco. Information, 11(12), 542.

[43] K. Mittal, D. Khanduja, and P. C. Tewari, "An insight into 'Decision Tree Analysis'"," World Wide Journal of Multidisciplinary Research and Development, vol. 3, no. 12, pp. 111–115, 2017.

[44] Kamble, S. J., & Kounte, M. R. (2020). Machine learning approach on traffic congestion monitoring system in internet of vehicles. Procedia Computer Science, 171, 2235-2241.

[45] Kashyap, A. A., Raviraj, S., Devarakonda, A., Nayak K, S. R., KV, S., & Bhat, S. J. (2022). Traffic flow prediction models–A review of deep learning techniques. Cogent Engineering, 9(1), 2010510.

[46] Knapińska, A., Lechowicz, P., & Walkowiak, K. (2021, June). Machine-learning based prediction of multiple types of network traffic. In International Conference on Computational Science (pp. 122-136). Springer, Cham.

[47] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), 3-24.

[48] Kumar, R., & Verma, R. (2012). Classification algorithms for data mining: A survey. International Journal of Innovations in Engineering and Technology (IJIET), 1(2), 7-14.

[49] Lee, S., & Fambro, D. B. (1999). Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transportation Research Record: Journal of the Transportation Research Board, 1678(1), 179–188.

[50] Li, J., Boonaert, J., Doniec, A., & Lozenguez, G. (2021). Multi-models machine learning methods for traffic flow estimation from Floating Car Data. Transportation Research Part C: Emerging Technologies, 132, 103389.

[51] Li, Y., & Jiang, W. (2019, November). Research on short-term Traffic flow Prediction Based on Big Data Environment. In 2019 Chinese Automation Congress (CAC) (pp. 1758-1762). IEEE.

[52] Lippi, M.; Bertini, M.; Frasconi, P. 2013. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning, IEEE Transactions on Intelligent Transportation Systems 14(2):871–882.

[53] Lizotte DJ, Laber EB (2016) Multi-Objective Markov Decision Processes for Data-Driven Decision Support. Journal of Machine Learning Research 17:211:1–211:28.

[54] M. Arslan, A.-M. Roxin, C. Cruz, and D. Ginhac, "A review on applications of big data for disaster management," in 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2017, pp. 370–375.

[55] M. C. Council. Open data portal of the madrid city council. [Online]. Available: (http://datos.madrid.es).

[56] Magalhaes, R. P., Lettich, F., Macedo, J. A., Nardini, F. M., Perego, R., Renso, C., & Trani, R. (2021). Speed prediction in large and dynamic traffic sensor networks. Information Systems, 98, 101444.

[57] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

[58] Makaba, T., Doorsamy, W., & Paul, B. S. (2020). Exploratory framework for analyzing road traffic accident data with validation on Gauteng province data. Cogent Engineering, 7(1), 1834659.

[59] Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784-2817.

[60] McLachlan, G. J. (2005). Discriminant analysis and statistical pattern recognition. John Wiley & Sons.

[61] Meena, G., Sharma, D., & Mahrishi, M. (2020, February). Traffic prediction for intelligent transportation system using machine learning. In 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE) (pp. 145-148). IEEE.

[62] Navarro-Espinoza, A., López-Bonilla, O. R., García-Guerrero, E. E., Tlelo-Cuautle, E., López-Mancilla, D., Hernández-Mejía, C., & Inzunza-González, E. (2022). Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms. Technologies, 10(1), 5.

[63] Nellore K, Hancke G (2016) A survey on urban traffic management system using wireless sensor networks. Sensors 16:157.

[64] Nie, L., Ning, Z., Obaidat, M. S., Sadoun, B., Wang, H., Li, S., ... & Wang, G. (2020). A reinforcement learning-based network traffic prediction mechanism in intelligent internet of things. IEEE Transactions on Industrial Informatics, 17(3), 2169-2180.

[65] Nikam, S. S. (2015). A comparative study of classification techniques in data mining algorithms. Oriental Journal of Computer Science and Technology, 8(1), 13-19.

[66] Obulesu, O., Mahendra, M., & ThrilokReddy, M. (2018, July). Machine learning techniques and tools: A survey. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 605-611). IEEE.

[67] Oladipo, I. D., AbdulRaheem, M., Awotunde, J. B., Bhoi, A. K., Adeniyi, E. A., & Abiodun, M. K. (2022). Machine Learning and Deep Learning Algorithms for Smart Cities: A Start-of-the-Art Review. IoT and IoE Driven Smart Cities, 143-162.

[68] Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." International Journal of Remote Sensing. 26 (1): 217–222.

[69] Patel, P., Narmawala, Z., & Thakkar, A. (2019). A survey on intelligent transportation system using internet of things. Emerging Research in Computing, Information, Communication and Applications, 231-240.

[70] Priyanka and D. Kumar, "Decision tree classifier: a detailed survey," International Journal of Information and Decision Sciences, vol. 12, no. 3, pp. 246–269, 2020.

[71] Qu, Z., & Li, J. (2022). Short-term Traffic Flow Forecast on Basis of PCA-Interval Type-2 Fuzzy System. In Journal of Physics: Conference Series (Vol. 2171, No. 1, p. 012051). IOP Publishing.

[72] Qureshi, K.N., & Abdullah, A.H. (2013). A Survey on Intelligent Transportation Systems. Middle-East Journal of Scientific Research, 15, 629-642.

[73] R. L. Villars, C. W. Olofson, and M. Eastwood, "Big data: What it is and why you should care," White paper, IDC, vol. 14, pp. 1–14, 2011.

[74] Rajkumar, S. C. (2019). Optimized traffic flow prediction based on cluster formation and reinforcement learning. International Journal of Communication Systems, e4178.

[75] Ramchandra, N. R., & Rajabhushanam, C. (2021, June). Traffic Prediction System Using Machine Learning Algorithms. In I3CAC 2021: Proceedings of the First International Conference on Computing, Communication and Control System, I3CAC 2021, 7-8 June 2021, Bharath University, Chennai, India (p. 424). European Alliance for Innovation.

[76] Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.

[77] Resende, P. A. A., & Drummond, A. C. (2018). A survey of random forest-based methods for intrusion detection systems. ACM Computing Surveys (CSUR), 51(3), 1-36.

[78] S. S. Gavankar and S. D. Sawarkar, "Eager decision tree", in 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, Apr. 2017, pp. 837–840.

[79] Sayed, S. A., Abdelhamid, Y., & Hefny, H. A. (2023). Traffic Flow Prediction Using Big Data and Geographic Information Systems: A Survey of Data Sources, Frameworks, Challenges, and Opportunities. Int. J. Com. Dig. Sys, 14(1).

[80] Shahraki, A., Taherkordi, A., & Haugen, Ø. (2021). TONTA: Trend-based online network traffic analysis in ad-hoc IoT networks. Computer Networks, 194, 108125.

[81] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. Nature 529(7587):484–489.

[82] Singh, G., Al'Aref, S. J., Van Assen, M., Kim, T. S., van Rosendael, A., Kolli, K. K., ... & Min, J. K. (2018). Machine learning in cardiac CT: basic concepts and contemporary data. Journal of Cardiovascular Computed Tomography, 12(3), 192-201.

[83] Steffen, T., & Lichtenberg, G. (2022, February). A Machine Learning Approach to Traffic Flow Prediction using CP Data Tensor Decompositions. In IFAC World Congress 2020. Loughborough Research Repository.

[84] Sun, P., Boukerche, A., & Tao, Y. (2020). SSGRU: A novel hybrid stacked GRU- based traffic volume prediction approach in a road network. Computer Communications, 160, 502–511.

[85] Suneel Kuamr. (2022). Traffic Flow Prediction Using Machine Learning Algorithms. International Research Journal of Engineering and Technology (IRJET), 9(4), 2995–3004.

[86] Świechowski, M., Godlewski, K., Sawicki, B., & Mańdziuk, J. (2021). Monte Carlo tree search: A review of recent modifications and applications. arXiv preprint arXiv:2103.04931.

[87] T. C. M. A. for Planning (CMAP). My daily travel survey. [Online]. Available: (https://www.cmap.illinois.gov/data/transportation/travel-survey#My_Daily_Travel_Survey).

[88] T. Kolajo, O. Daramola, and A. Adebiyi, "Big data stream analysis: a systematic literature review," Journal of Big Data, vol. 6, no. 1, p. 47, 2019.

[89] T. S. D. Center. 2010–2012 california household travel survey. [Online]. Available: (https://www.nrel.gov/transportation/secure-transportation-data/tsdc-california-travel-survey.html).

[90] Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. 2019 International Conference on Intelligent Computing and Control Systems (ICCS).

[91] Tianchi. Kdd cup 2017 highway tollgates traffic flow prediction dataset. [Online]. Available: (https://tianchi.aliyun.com/dataset/dataDetail?dataId=60).

[92] Tin Kam Ho. 1995. Random decision forests. In 3rd International Conference on Document Analysis and Recognition - Volume 1 (ICDAR'95). IEEE Computer Society, 278–282.

[93] Tin Kam Ho. 1998. The random subspace method for constructing decision forests. 20, 8 (Aug. 1998), 832–844.

[94] Tu, M., Liu, B., & Zhong, F. (2019, December). Research of intelligent traffic flow prediction algorithm. In Journal of Physics: Conference Series (Vol. 1423, No. 1, p. 012038). IOP Publishing.

[95] Upadhyaya, S., & Mehrotra, D. (2022). The Facets of Machine Learning in Lane Change Prediction of Vehicular Traffic Flow. In Proceedings of International Conference on Intelligent Cyber-Physical Systems (pp. 353-365). Springer, Singapore.

[96] Van Der Voort, M., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. Transportation Research Part C: Emerging Technologies, 4(5), 307–318.

[97] Veropoulos, K., Campbell, C. & Cristianini, N. (1999). Controlling the Sensitivity of Support Vector Machines. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI99).

[98] W. Jiang and J. Luo, "Big data for traffic estimation and prediction: a survey of data and tools," Applied System Innovation, vol. 5, no. 1, p. 23, 2022.

[99] W. Jiang and J. Luo, “Big data for traffic estimation and prediction: a survey of data and tools,” Applied System Innovation, vol. 5, no. 1, p. 23, 2022.

[100] Wang, J., Pradhan, M. R., & Gunasekaran, N. (2022). Machine learning-based human-robot interaction in ITS. Information Processing & Management, 59(1), 102750.

[101] Wang, Z., Chu, R., Zhang, M., Wang, X., & Luan, S. (2020). An improved hybrid highway traffic flow prediction model based on machine learning. Sustainability, 12(20), 8298.

[102] Watkin, C. J. C. H., & Dayan, P. (1992). Technical note Q-learning. Machine Learning, 8(3), 279-292.

[103] Watkins, C.J.C.H. (1989). Learning from delayed rewards. Ph.D. Thesis, University of Cambridge, England.

[104] Weerasekera, R., Sridharan, M., & Ranjitkar, P. (2020). Implications of spatiotemporal data aggregation on short-term traffic prediction using machine learning algorithms. Journal of Advanced Transportation, 2020.

[105] Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. Transportation Research Record: Journal of the Transportation Research Board, 1776(1), 194–200.

[106] Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of Transportation Engineering, 129(6), 664–672.

[107] World Health Organization. (2018). “GLOBAL STATUS REPORT ON ROAD SAFETY 2018 SUMMARY.” [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/277370/WHO-NMH-NVI-18.20-eng.pdf?ua=1.

[108] X. Shi, Z. Gao, L. Lausen, H. Wang, D.Y. Yeung, W.K. Wong, and W.C. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," Advances in Neural Information Processing Systems," pp. 5622–5632, 2017.

[109] Xiao, J., Xiao, Z., Wang, D., Bai, J., Havyarimana, V., & Zeng, F. (2019). Short-term traffic volume prediction by ensemble learning in concept drifting environments. Knowledge-Based Systems, 164, 213-225.

[110] Y. An, Y. Wang, and H. Meng, "Multi-task deep learning for user intention understanding in speech interaction systems," 2017.

[111] Y. Zhao and Y. Zhang, “Comparison of decision tree methods for finding active objects,” Advances in Space Research, vol. 41, no. 12, pp. 1955–1959, 2008.

[112] Yali Amit and Donald Geman. 1997. Shape quantization and recognition with randomized trees. Neural Computation 9, 7 (1997), 1545–1588.

[113] Yang H, Dillon TS, Chang E et al (2019) Optimized configuration of exponential smoothing and extreme learning machine for traffic flow forecasting[J]. IEEE Trans Ind Inf 15(1):23–34.

[114] Zahid, M., Chen, Y., Jamal, A., & Memon, M. Q. (2020). Short term traffic state prediction via hyperparameter optimization based classifiers. Sensors, 20(3), 685.

[115] Zeroual, A., Harrou, F., & Sun, Y. (2021, December). Predicting road traffic density using a machine learning-driven approach. In 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.

[116] Zheng, L., Yang, J., Chen, L., Sun, D., & Liu, W. (2020). Dynamic spatial-temporal feature optimization with ERI big data for Short-term traffic flow prediction. Neurocomputing, 412, 339-350.

[117] Zou, W., & Xia, Y. (2019). Back propagation bidirectional extreme learning machine for traffic flow time series prediction. Neural Computing and Applications, 31(11), 7401-7414.